DOI：10.11931/guihaia.gxzw202304069

珍稀濒危植物越南小花金花茶的叶绿体基因组特征分析

邓永彪，张进，蓝伦礼，赵博*

(桂林医学院 药学院生药学教研室，广西 桂林 541199)

摘要：越南小花金花茶(*Camellia minima*) 是一种珍稀濒危的金花茶组植物，目前尚未有其叶绿体基因组的相关研究。该研究利用 Illumina HiSeq 2000 平台对越南小花金花茶进行了叶绿体基因组序列的测序、组装、注释和分析。结果表明：(1) 越南小花金花茶叶绿体基因组全长 156 961 bp，为典型的四分体结构，共注释到 136 个基因，其中包含 87 个蛋白编码基因，41 个 tRNA 基因和 8 个 rRNA 基因。(2) 分析鉴定出 66 个 SSR 位点和 39 个重复序列。(3) 密码子偏好使用以 A/U 结尾的密码子，综合 ENC 绘图、PR2-plot 和中性分析推测自然选择是影响密码子使用模式的主导因素。(4) 边界分析显示 *ycf*1 基因的长度和位置在不同金花茶组植物间存在差异。(5) 对已发表的金花茶组植物叶绿体基因组的系统发育分析发现越南小花金花茶与小花金花茶聚为一支，亲缘关系很近。该研究结果为探索物种进化和提高外源基因表达提供了重要的参考信息，同时为后续的金花茶组植物的保护和利用提供了理论基础。

关键词：越南小花金花茶，叶绿体基因组，特征分析，系统发育分析，密码子偏好性

中图分类号：Q943　　　　**文献标识码：**A　　　　**文章编号：**

Analysis of chloroplast genome features of endangered and rare plant *Camellia minima*

DENG Yongbiao, ZHANG Jin, LAN Lunli, ZHAO Bo*

(*Department of Pharmacognosy, Pharmacy School, Guilin Medical University*, Guilin 541199, Guangxi, China)

Abstract: *Camellia minima*, a rare and endangered species of Sect. *Chrysantha*, has not been previously explored in terms of its chloroplast genome. Utilizing the Illumina HiSeq 2000 platform, the chloroplast genome sequence of *C. minima* was sequenced, assembled, annotated, and analysed, The results showed that: (1) The chloroplast genome of *C. minima* was 156 961 bp in length, embodies a typical tetrad structure, and contains 136 annotated genes, including 87 protein-coding genes, 41 tRNA genes, and 8 rRNA genes. (2) The analysis identified 66 SSR loci and 39 repetitive sequences. (3) Codons prefer to use codons ending in A/U. Comprehensive ENC mapping, PR2-plot, and neutral analyses suggest natural selection as a primary factor shaping codon usage patterns. (4) Boundary analysis indicated variation in the length and position of the *ycf*1 gene among different species of yellow *Camellia*. (5) Phylogenetic analysis of the chloroplast genomes of published species of Sect. *Chrysantha* revealed that *C. minima* was most closely

基金项目：国家自然科学基金项目（32060090)；广西科技基地和人才专项（桂科 AD19110089)，广西自然科学基金项目（2021JJA130119）[Supported by National Natural Science Foundation of China (32060090); Guangxi Science and Technology Base and Talent Special Project (GUI Ke AD19110089); Natural Science Foundation of Guangxi (2021JJA130119).]。

作者简介：邓永彪(2000-)，硕士研究生，研究方向为中药资源，(E-mail)dengyongbiao1@gmail.com。

***通讯作者：**赵博(1981-)，博士，教授，研究方向为药用植物资源学，(E-mai)122017017@glmc.edu.cn。

related to *C. micrantha.* This study provides crucial insights for exploring species evolution and enhancing exogenous gene expression, establishing a theoretical foundation for the conservation and utilization of species of Sect. *Chrysantha* in the future.
Key words: *Camellia minima*, chloroplast genome, characteristic analysis, phylogenetic analysis, codon preference

金花茶组(Sect. *Chrysantha*)隶属于山茶科(Theaceae)山茶属(*Camellia* L)，主要分布在我国的广西、云南以及越南部分地区(赛璇，2018)。其生长环境为人类活动踪迹较少的喀斯特地貌和潮湿山地等(李桂娥等，2022)。所有金花茶组植物都被列入我国国家重点保护野生植物名录，保护等级为二级(国家林业和草原局，2021)。金花茶组植物被称为“植物界的大熊猫”、“茶中皇后”，因为该组植物是山茶属唯一拥有金黄色花瓣的类群(张蕾等，2018，吴丽君等，2020)。金花茶组植物的花瓣和叶片中含有丰富的黄酮，常被用作保健食品与饮品(刘青等，2021)。金花茶植物由于频繁的种间杂交和多倍化导致其分类研究十分困难(Zhang et al., 2019)。Wei 等(2022)基于双酶切位点相关的简化基因组测序(ddRAD)、转录组和核糖体内部转录间隔区(nrITS)等数据对 20 个中国金花茶物种进行系统发育重建，结合形态学证据确定了物种之间的亲缘关系。研究表明，强烈的杂交/渗入信号表明网状进化是造成核基因数据和叶绿体基因组数据构建的系统发育树不一致的主要原因。该研究仅采用了片段较小的叶绿体基因组的 SSC 区，更多完整的叶绿体基因组数据的积累将为今后金花茶属植物的系统发育和网状进化关系的研究提供更多的分子证据，也可以为金花茶组植物的保护与综合利用提供基础理论支持。

四分体结构是叶绿体基因组的典型特征，由一个大单拷贝(large single-copy, LSC)区、一个小单拷贝(small single-copy, SSC)区和两个倒置重复序列(inverted repeat, IRs)显示和编码 110~130 个基因，大小范围为 120~180 kb(Li et al., 2021)。相较于核基因组，叶绿体基因组在基因结构、基因内容、基因序列方面更加稳定，较慢的进化速率，因此被广泛应用于系统发育、DNA 条形码、基因工程和种群间亲缘关系等研究中(Dong et al., 2018)。近年来叶绿体基因组测序的成本大大降低，越来越多的叶绿体基因组数据被成功应用于植物系统发育和进化研究。同时叶绿体基因组中包含大量的重复序列是研究物种进化进程以及遗传特征的重要依据(Hui et al., 2014)，其中简单重复序列(SSRs)又称微卫星序列可以作为有效的分子标记来检测种群多态性，广泛应用于分子辅助育种和物种保护等方面 (Cavalier, 2002)。密码子是连接核酸和蛋白质的纽带，在遗传信息的传递过程中起着重要的作用(Liu et al., 2012)。研究物种密码子使用模式并确定最优密码子，这有助于设计基因表达载体来提高目的基因的表达量，在品种改良方面具有重要应用价值(Qi et al., 2015, 胡晓艳等，2019)。

越南小花金花茶(*C. minima*)为金花茶组植物之一，可做观赏植物和嫁接砧木(George & Anthony, 2015)。目前尚未有关于越南小花金花茶叶绿体基因组的相关研究，这限制了对其遗传特征和系统发育关系的理解。为此，本研究利用高通量测序技术对越南小花金花茶进行了叶绿体全基因组测序组装并探讨了以下科学问题：(1) 表征了越南小花金花茶叶绿体基因组图谱和序列特征；（2）分析了越南小花金花茶叶绿体基因组密码子使用的偏好性并推测影响密码子使用模式主导因素；(3)探究了越南小花金花茶与其近缘物种在 IR 和 SC 边界区域碱基分布的差异；(4)明确了越南小花金花茶在金花茶组植物系统发育中的位置。本研究不仅为后续开展越南小花金花茶物种鉴定、遗传多样性分析、叶绿体基因工程以及分子育种等研究提供了重要的理论基础，同时也为金花茶组植物的系统发育研究提供了更丰富的叶绿体基因组数据。

1 材料和方法

1.1 植物材料

越南小花金花茶采集于中国广西壮族自治区南宁市金花茶公园(108°21′43" E 、22°49′30″ N)，采集单株新鲜幼嫩的叶片，然后装进有硅胶的自封袋内低温保存，用于后续提取总 DNA，标本保存在广西壮族自治区中国科学院广西植物研究所标本馆，标本凭证为：(CSF2020003)。

1.2 DNA 提取和测序

使用改良的 CTAB 法从越南小花金花茶干燥的叶片中提取总 DNA，叶绿体基因组测序服务由北京格致博雅生物科技有限公司提供。使用 fastp 软件处理测序获得的原始数据，除去过多的 N 序列、接头序列和过短序列，输出大小为 1 005.18Mb 的 clean data，用于后续的叶绿体基因组组装(Gu et al., 2018)。

1.3 基因组组装注释和序列表征

使用 getOrganelle v1.7.6.1 组装 clean data，参考数据库为 em bplant_pt(陆生植物叶绿体)，最大扩充循环数为 20，调用 Spedas 的 K-mer 值为 21, 45, 65, 85, 105, 127；拼接完成后用 Bandage 软件可视化(Jian et al., 2018)。使用叶绿体基因组在线注释网站 CPGAVAS2 (http://47.96.249.172:16019/analyzer/home)，以 *C. parvipetala* (NC_067089.1)为参照基因组进行注释。使用 GB2sequin 检查序列分区是否存在颠倒并手动调整准确位置，得到完整的注释结果(Shi et al., 2019, Pascal & Stephan, 2018)。最后使用在线网站 CPGview (http://www.1kmpg.cn/cpgview/)绘制了叶绿体基因组圈图(Liu et al., 2023)。注释完成的叶绿体基因组提交至 GenBank 数据库，登录号为 NC_069310.1，相应的 SRA、BioProject 和 BioSample 编号分别为 SRR20648317、PRJNA861872 和 SAMN29930849。

1.4 密码子使用偏好性分析

1.4.1 密码子相关参数计算

在剔除重复基因、含有终止密码子和长度小于 300 bp 的编码序列(Coding sequence, CDS)后，选取 52 条 CDS 序列进行密码子分析。使用 Codon W 1.4.2 软件对相对同义密码子使用度(Relative synonymous codon usage，RSCU)进行分析。利用 EMBOSS 工具包中的 CUSP 程序分析总 GC 含量以及第 1、第 2 和第 3 个碱基位置(GC1、GC2、GC3)的 GC 含量，CHIPS 程序用于分析有效密码子数(ENC)(朱伟垚等，2022)。最后采用 R 软件中的 cor()函数计算相关性。

1.4.2 密码子中性绘图、ENc-plot 及 PR2-plot 绘图

在中性绘图分析中，以 GC12 和 GC3 做散点图，研究密码子 3 个位置碱基之间的相关性，从而分析突变压力和自然选择对于密码子偏好性的影响(Wei et al., 2014)。GC12 为 GC1 和 GC2 的平均值。如果 GC12 和 GC3 之间显著相关，即 R^2 越大且趋于 1 时，GC12 与 GC3 的相关程度越高，说明突变压力是密码子使用偏向的决定因素。相反，如果相关性不显著，则回归曲线斜率斜率偏低甚至接近于零，表明密码子偏好性由自然选择主导(Sueoka, 1988)。

ENC-plot 绘图进一步分析了密码子偏好性受碱基组成对的影响，以 GC_3 为横坐标、有效密码子数(ENC ）为纵坐标进行绘制散点图，根据下式计算：$ENC = 2 + GC_3 + 29/[GC_3^2 + (1-GC_3)^2]$ (杨祥燕等，2021，辛雅萱等，2021)。当密码子偏好性仅受突变影响时，基因将沿着或接近标准曲线分布，而当基因落在标准曲线以下时则说明自然选择等因素是影响密码子偏好性的主要因素(Chakraborty et al., 2020)。

奇偶偏好性分析(Parity rule 2 plot, PR2-plot)用于计算各密码子第 3 位 A、T、C、G 的含量，通常用于估计突变压力和自然选择对密码子偏好性的影响(Xiang et al., 2015)。以 A3 /(A3 + T3)为 *y* 轴，G3 /(G3 + C3)为 *x* 轴绘制 PR2-plot 图。中心点(A = T, G = C)意味着自然

选择和突变压力之间没有偏差。如果基因均匀分布在中心点周围，则认为密码子偏好性可能完全是由突变压力造成的，否则，密码子的使用可能受到自然选择和其他因素的影响(Xiang et al., 2015)。

1.5 重复序列分析

使用在线网站 REPuter 对散在重复序列进行了计算。参数设置为：最大计算重复次数=200; 最小重复序列>30 bp；重复序列相似度>90%；汉明距离(Hamming Distance)= 3 (Kurtz et al., 2001)。简单重复序列使用在线网站 MISA 进行鉴定和统计分析。各重复单元(unit size)对应的的最少重复次数分别为:1～10、2～5、3～4、4～3、5～3、6～3，相邻 SSR 之间的最小单位间隔距离设置为 100 bp (Beier et al., 2017)。

1.6 叶绿体基因组 IR 边界分析

基于系统发育分析的结果，以簇蕊金花茶(*C. fascicularis*)叶绿体基因组为参考，使用本地部署的 IRscope 对越南小花金花茶及其近缘物种叶绿体基因组的 LSC/IRb/SSC/IRa 边界进行可视化分析(Amiryousefi et al., 2018)。

1.7 系统发育分析

将测序组装得到的越南小花金花茶以及从 NCBI 数据库中获得的 23 个为金花茶组植物的叶绿体基因组全长序列，以 *Polyspora penangensis* 为外类群，构建系统发育树。具体流程：使用 MAFFT(version 7.505)进行多序列比对，使用 trimAI(V1.4)去除多序列比对结果中的低质量区域，以提高比对结果的质量和准确性。使用 IQ-TREE2 软件在 TVM+F+I+I+R6 模型下，以最大似然法(Maximum Likelihood, ML)构建系统发育树(Katoh & Standley, 2003; Minh et al., 2021)。

2 结果

2.1 叶绿体基因组特征

越南小花金花茶叶绿体基因组为环状双链四分体结构(图 1)，长度为 156 961 bp，其中有两个长度为 26 047 bp 的反向重复(IR)区域被一个 86 600 bp 的大单拷贝(LSC)区域和一个 18 267 bp 的小单拷贝(SSC)区域分开。叶绿体基因组的 GC 总含量为 37.32%，其中 LSC 区为 35.33%，SSC 区为 30.60%，IR 区为 42.98%。

越南小花金花茶叶绿体基因组包含 136 个功能基因(表 1)，包括 87 个蛋白质编码基因，41 个 tRNA 基因和 8 个 rRNA 基因，在这些基因中，与自我复制相关的基因有 78 个，与光合作用有关的基因有 45 个，其他功能未知的基因共有 13 个。基因组中有 16 个双拷贝基因(*ndh*B, *rpl*2, *rpl*23, *rps*7, *rps*12, *rrn*4.5S, *rrn*5S, *rrn*16S, *rrn*23S, *trn*I-CAU, *trn*L-CAA, *trn*N-GUU, *trn*R-ACG, *trn*V-GAC, *ycf*2, *ycf*15), 2 个四拷贝基因(*trn*I-GAU, *trn*A-UGC)。在越南小花金花茶的叶绿体基因组中共有 10 个基因含有内含子这些基因分别是, petB, petD, *rps*16, *rpl16, rpl12, atp*F, *rpo*C1, *clp*P, *ndh*B, *ndh*A，其中 *rpl*12 和 *ndh*B 基因都含有 2 个内含子。

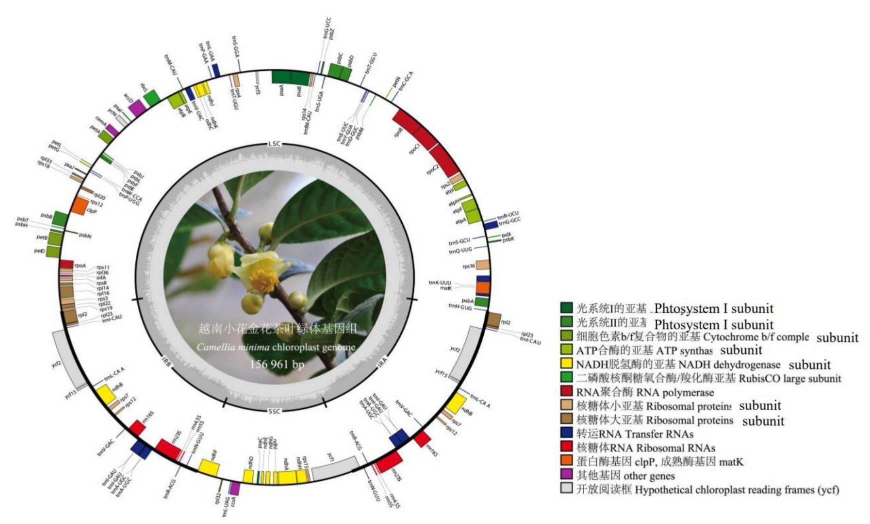


图 1 越南小花金花茶的叶绿体基因组圈图

Fig.1 Chloroplast genome physical map of *Camellia minima*

表 1 越南小花金花茶叶绿体基因组的注释基因信息

Table 1 Annotated gene information of *Camellia minima* chloroplast genome

功能 Function	家族 Family name	基因名称 Gene name
自我复制 self-replication	核糖体小亚基 Ribosomal proteins (SSU)	*rps2, rps3 , rps4, rps8, rps11, rps14, rps15, rps16, rps18, rps19, rps7*, rps12**
	核糖体 RNA rRNA	*rrn4.5S*, rrn5S*, rrn16S*, rrn23S**
	核糖体大亚基 Ribosomal proteins (LSU)	*rpl14, rpl16, rpl20, rpl22, rpl32, rpl33, rpl36, rpl2*, , rpl23**
	DNA 依赖性 RNA 聚合酶 DNA-dependent RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*
	转运 RNA tRNA	*trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU, trnG-GCC, trnG-UCC, trnH-GUG, trnK-UUU, trnL-UAA, trnL-UAG, trnM-CAU, trnP-UGG, trnQ-UUG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-UAC, trnW-CCA, trnY-GUA*, trnR-ACG*, trnL-CAA*, trnV-GAC*, trnN-GUU*, trnI-CAU*, *rnA-UGC**, trnI-GAU**,*
光合作用 Photosynthesis	ATP 合酶亚基 ATP synthase subunit	*atpA, atpB, atpE, atpF, atpH, atpI*

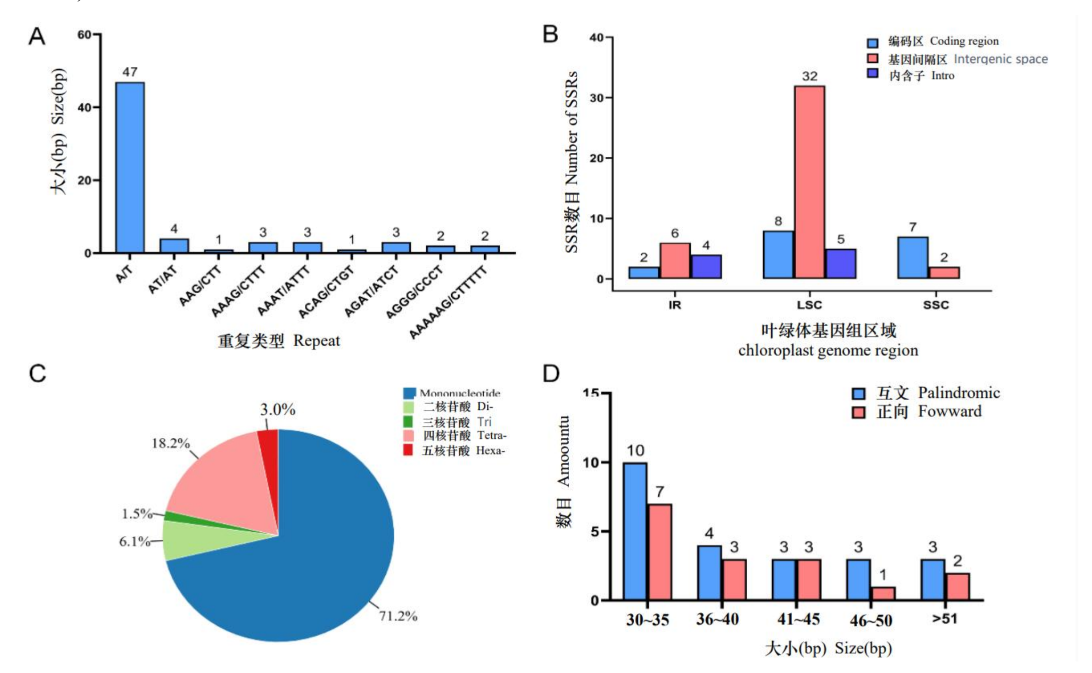
	NADH－脱氢酶的亚基 NADH-dehydrogenase subunits	*ndhA, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK, ndhB**
	细胞色素 b/f 复合物的亚基 Subunits of cytochrome b / f complex	*petA, petB, petD, petG, petL, petN*
	光系统Ⅰ的亚基 Subunits of photosystem I	*psaA, psaB, psaC, psaI, psaJ*
	光系统Ⅱ的亚基 Subunits of photosystem Ⅱ	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
	二磷酸核酮糖氧合酶	*rbcL*
其他 other	假定叶绿体阅读框 Hypothetical chloroplast reading frames (ycf)	*ycf1, ycf3, ycf4, ycf15*, ycf2**
	乙酰－CoA－羧化酶的亚基 Subunits of acetyl-CoA-carboxylase	*accD*
	包膜蛋白基因 Envelope protein gene	*cemA*
	c 型细胞色素合成基因 Type c cytochrome synthesis gene	*ccsA*
	蛋白酶基因 Protease gene	*clpP*
	成熟酶基因 Mature enzyme gene	*matK*
	转录起始因子 Transcription initiation factor	*infA*

注：*表示存在一个内含子;**表示存在两个内含子。

Note:* indicates the presence of one intron; ** indicates the presence of two introns.

2.2 重复序列分析

通过 MISA 在线网站分析，发现越南小花金花茶的叶绿体基因组中含有的 SSR 位点数目为 66 个，类型包括单核苷酸、二核苷酸、三核苷酸、四核苷酸和六核苷酸，且大多数为单核苷酸重复(图 2：C)。这些类型的 SSR 有 17 个分布在蛋白编码区、40 个分布在基因间隔区、9 个分布在内含子区（图 2：B）。通过 REPuter 软件分析了越南小花金花茶叶绿体基因组中的散在重复序列，共检测到 39 个散在重复序列，分别是 16 个正向重复序列(forword)和 23 个回文重复序列(palindromic)，但没有检测到反向(reverse)和互补(complement)重复序列的存在。这些散在重复序列的长度在 30~64 bp 之间，且大部分位于 IR 区的 *ycf*2 基因中(图 2：D)。



A. SSR 的类型与数量；**B**. LSC、SSC 和 IR 区域的蛋白编码区、基因间的间隔区和内含子中 SSR 的数量；**C**. 不同类型 SSR 的占比；**D**. 散在重复序列类型与数量。

A. Type and number of SSRs; **B**. Presence of SSRs in the proteincoding regions, intergenic spaces (IGSs) and introns of LSC, SSC and IR regions; **C**. Proportion of different types of SSRs; **D**. Type and number of long repeats.

图 2 越南小花金花茶叶绿体基因组的重复序列分布

Fig. 2 Distribution of of different types of SSRs

2.3 密码子偏好性分析

2.3.1 同义密码子相对使用度分析

对越南小花金花茶叶绿体基因组的 52 个大于 300 bp 的蛋白质编码序列进行密码子偏好性分析发现，其中相对同义密码子使用度(RSCU)值大于 1 的密码子有 30 个，其中 13 个以 A (30 中 13)或 U (30 中 16)结尾，以 G 结尾的数目仅有 1 个；RSCU 值小于 1 的 32 个密码子多以 C (32 个中 16 个)或 G (32 中 13)结尾 ，表明密码子偏好使用以 A/U 结尾的密码子(图 3)。

RSCU

苯丙氨酸 Phe	UUU 1.20	UUC 0.80	0.00	0.00	0.00	0.00
亮氨酸 Leu	UUA 1.36	CUU 1.21	UUG 1.21	CUA 0.91	CUC 0.76	CUG 0.55
异亮氨酸 Ile	AUU 1.22	AUA 1.01	AUC 0.77	0.00	0.00	0.00
甲硫氨酸 Met	AUG 1.00	0.00	0.00	0.00	0.00	0.00
缬氨酸 Val	GUU 1.34	GUA 1.26	GUC 0.73	GUG 0.67	0.00	0.00
酪氨酸 Tyr	UAU 1.33	UAC 0.67	0.00	0.00	0.00	0.00
终止子 TER	UAA 1.21	UGA 1.00	UAG 0.79	0.00	0.00	0.00
组氨酸 His	CAU 1.40	CAC 0.60	0.00	0.00	0.00	0.00
谷氨酰胺 Gln	CAA 1.38	CAG 0.62	0.00	0.00	0.00	0.00
天冬酰胺 Asn	AAU 1.39	AAC 0.61	0.00	0.00	0.00	0.00
离氨酸 Lys	AAA 1.38	AAG 0.62	0.00	0.00	0.00	0.00
天冬氨酸 Asp	GAU 1.42	GAC 0.58	0.00	0.00	0.00	0.00
谷氨酸 Glu	GAA 1.37	GAG 0.63	0.00	0.00	0.00	0.00
丝氨酸 Ser	UCU 1.54	UCA 1.33	UCC 1.12	AGU 0.79	UCG 0.68	AGC 0.54
脯氨酸 Pro	CCA 1.27	CCU 1.09	CCG 0.63	0.00	0.00	0.00
苏氨酸 Thr	ACA 1.25	ACU 1.14	ACC 0.99	ACG 0.62	0.00	0.00
丙氨酸 Ala	GCU 1.28	GCA 1.19	GCC 0.91	GCG 0.62	0.00	0.00
半胱氨酸 Cys	UGU 1.20	UGC 0.80	0.00	0.00	0.00	0.00
色氨酸 Trp	UGG 1.00	0.00	0.00	0.00	0.00	0.00
精氨酸 Arg	AGA 2.11	AGG 1.10	CGA 1.01	CGG 0.69	CGU 0.64	CGC 0.45
甘氨酸 Gly	GGA 1.48	GGU 0.97	GGG 0.95	GGC 0.60	0.00	0.00

2.00

0.00

图中不同的颜色表示 RSCU 值的大小，随着颜色变深密码子使用率增加。

Different colors in the figure represent the size of the RSCU value. As the color darkens, the codon usage rate increases.

图 3 越南小花金花茶的 RSCU 值热图

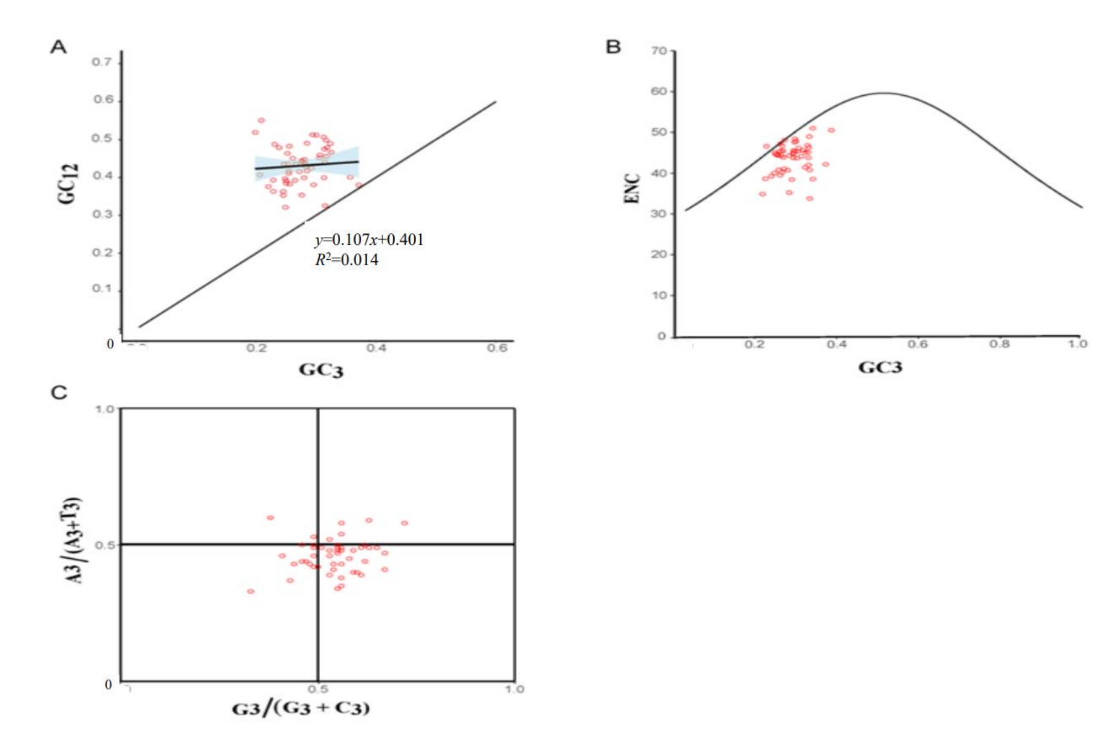
Fig. 3 RSCU heat map of *Camellia minima*

2.3.2 密码子中性绘图、ENc-plot 及 PR2-plot 分析

对越南小花金花茶叶绿体基因进行中性图(GC12 vs GC3)分析(图 4：A)。结果显示叶绿体基因分布相对集中，但并不密集。GC12 和 GC3 的平均值分别为 43.10%和 27.49%。GC12 和 GC3 的相关系数为 r=0.118(R^2=0.014)，回归曲线斜率为 0.107，表明 GC12 与 GC3 无显著相关性。由此推断密码子使用偏好性受突变压力影响较小，而自然选择等因素的贡献更大。

为了评估越南金花茶叶绿体基因组中蛋白编码基因的密码子偏好程度，计算并分析了 ENC 值。如图 4B 所示，大部分基因的 ENC 值低于预期值，位于标准曲线下方，表明密码子使用偏好性主要受自然选择等因素的影响，而突变压力仅起次要作用。为了进一步反映差异，分析了越南小花金花茶叶绿体基因的 ENC 频率分布(表 2)。ENC 比值在-0.05~0.45 之间，其中 9 个(17.00%)基因分布在-0.05~0.05 范围内，表明这部分基因分布在标准曲线周围，而剩余的大部分基因都距离标准曲线较远，以上结果说明了自然选择对越南小花金花茶叶绿体基因组密码子偏好的影响更大。

如果密码子偏好性完全受突变影响，则密码子第 3 位核苷酸 A、T、C、G 的使用频率相同。在本研究中，PR2 平面的 4 个区域并不是均匀分布的(图 4：C)。大部分基因位于平面的下半部分，尤其是右下方。因此，从碱基的使用频率来看，T>A，G>C，表明叶绿体基因组中 A/T 和 G/C 的密码子使用偏好性不平衡，推断越南小花金花茶叶绿体基因组的密码子偏好性同时受到突变压力和自然选择等因素的影响。



A. 中性绘图分析；**B.** ENC-GC3 关联分析；**C.** PR2-plot 关联分析。

A. Neutrality plot analysis; **B.** Analysis on ENC and GC3 relationship; **C.** Analysis on PR2-plot relationship.

图 4 越南小花金花茶叶绿体基因的进化因素分析

Fig. 4 Analysis of evolutionary forces of chloroplast genes in *Camellia minima*

表 2 ENC 比值频数分布

Table 2 Distribution of ENC ratio

组限 Class boundary	中值 Class mid value	频数 Number	频率 Frequency
-0.05~0.05	0	9	0.17
0.05~0.15	0.1	30	0.58
0.15~0.25	0.2	10	0.19
0.25~0.35	0.3	2	0.04
0.35~0.45	0.4	1	0.02

2.4 IR 边界分析

叶绿体基因组在 LSC、IRb、IRa 和 SSC 之间有四个边界：LSC/IRb 边界，IRb/SSC 边界，SSC/IRa 边界和 IRa/LSC 边界。本研究对越南小花金花茶及 6 个近缘金花茶组植物的叶绿体基因组边界进行了比较分析(图 5)。结果表明，在所有比较的叶绿体基因组中，IR 区仅有轻微的扩张和收缩，长度从 25 996 bp(四季花金花茶)到 26 096 bp(毛瓣金花茶)。在 LSC/IRb 边界处，所有物种的 *rps*19 基因都跨越了 LSC/IRb 边界，向 IRb 区域延伸了 46 bp。在 IRb/SSC 边界处，7 个金花茶组植物的边界均位于 *ycf*1 假基因和 *ndh*F 基因的重叠处，在越南金花茶和毛瓣金花茶中，*ndhF* 基因跨越 IRb/SSC 边界，分别有 39、25 bp 位于 IRb 区域中。在 SSC/IRa 边界处，*ycf*1 基因跨越该边界，包含在 IRa 区域内 1034~1055 bp。在 IRa/LSC 边界处，所有物种的边界均位于 *rps*19 拷贝基因和 *trn*H 基因之间，*trn*H 距离 IRa/LSC 边界 1 bp。

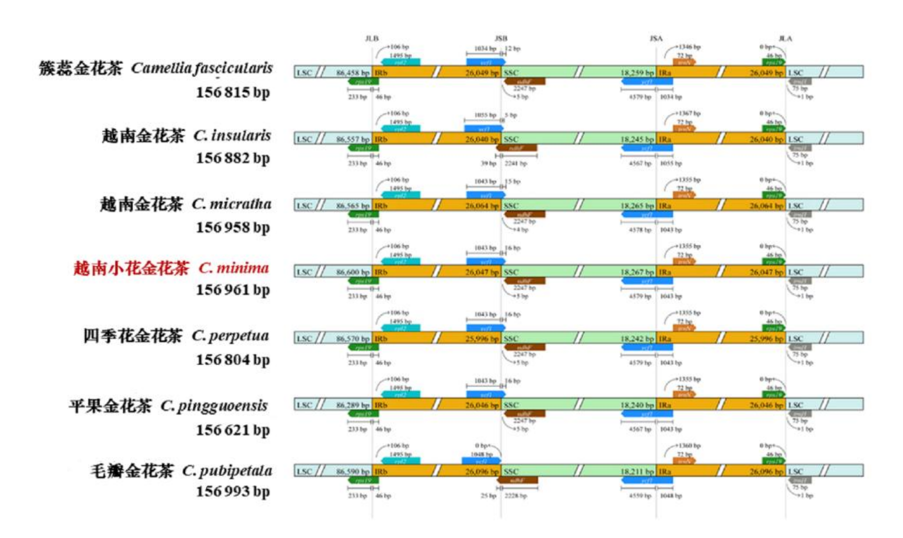
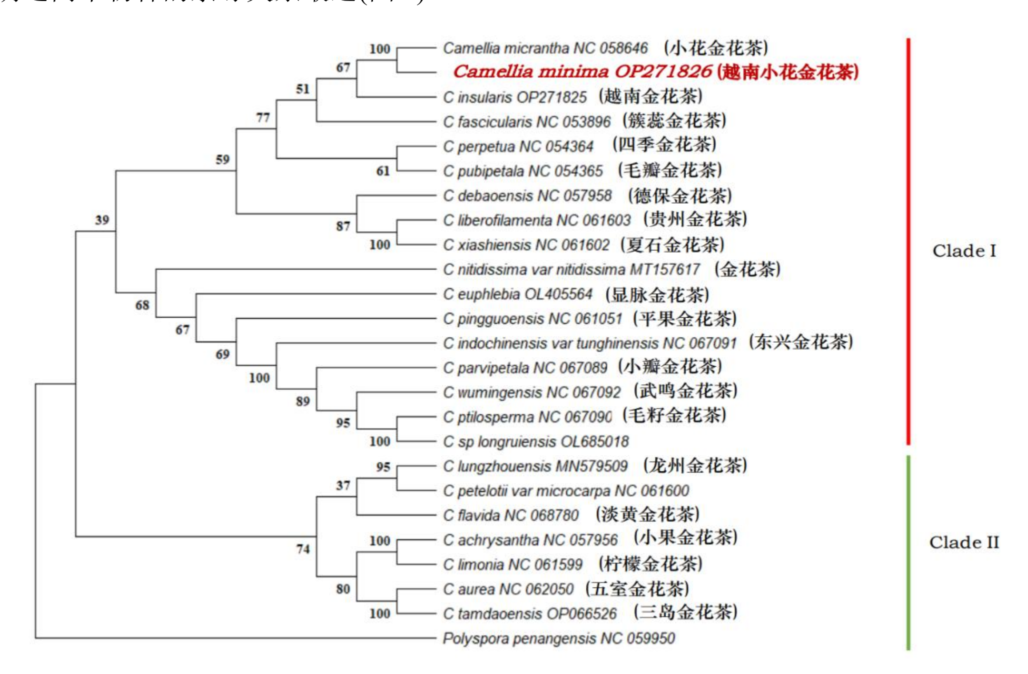


图 5 来自 7 种金花茶组植物的叶绿体基因组 IR 区域的边界分析

Fig. 5 Boundary analysis of IR regions of chloroplast genomes from seven yellow *camellia* species

2.5 系统发育分析

基于 25 种山茶科植物(包括 24 种金花茶组植物和 1 种大头茶属物种)叶绿体全基因组序列构建系统发育树，以 *P. penangensis* 作为外类群，利用最大似然法构建系统发育树。结果表明 24 种金花茶组植物主要分为 2 个分支 CladeI 分支和 CladeII 分支。CladeI 分支中，越南小花金花茶(*C. minima*)与小花金花茶(*C. micrantha*)以 100%的支持率聚为了一个单支，表明这两个物种的亲缘关系最近(图 6)。



支上的数字为 1 000 次自展支持率(%)。

The 1 000 bootstrap value (%) was shown above the branch.

图 6 基于 25 个叶绿体全基因组的系统发育树

Fig. 6 Phylogenetic tree based on 25 complete chloroplast genomes

3 讨论与结论

金花茶组植物均为国家二级保护植物(国家林业和草原局，2021)，具有较高的观赏和药用价值。其中越南小花金花茶是一种原产于越南北部的金花茶组植物，生长于潮湿荫蔽的山谷。由于其拥有较强的适应能力，被视为优良的嫁接砧木，近年来受到了园艺界的广泛关注(李桂娥等，2022)。越南小花金花茶叶绿体基因组为典型的环状双链四分体结构、长度为 156 961 bp，GC 含量为 37.32%，叶绿体全基因组注释到了 136 个基因，包括 87 个蛋白编码基因，41 个转运 RNA 基因和 8 个核糖体 RNA 基因, 共有 10 个基因含有内含子，其中 *rpl*12 和 *ndh*B 基因含有 2 个内含子，这与金花茶组其他植物的特征一致，可能与叶绿体基因组的特殊结构和复制方式有关(Henry et al., 2016; Wu et al., 2021)。此外，越南小花金花茶的叶绿体基因组长度、基因的类型和排列顺序、GC 含量等与山茶属的其他已发表物种 (例如山茶、油茶、凹脉金花茶等) 相似。这表明越南小花金花茶叶绿体基因组的进化可能是保守而缓慢的(丁祥青等，2022; Sophiarani et al., 2019)。

简单序列重复序列(SSRs)是由 DNA 链的滑移产生的，在基因组和叶绿体基因组中都有较广分布。与其他中性 DNA 区域相比，SSRs 通常具有更高的突变率。由于其具有非重组、单倍体和单亲遗传的特性，因此常被用作遗传标记，为植物群体遗传学和生态学及进化研究提供有价值的信息(Gui et al, 2020, Aii et al., 1997)。该研究在越南小花金花茶叶绿体基因组中共鉴定了 66 个 SSR 位点，主要位于大单拷贝区(LSC)，其中所有的单核苷酸重复都是由 A/T 组成，类似的结果在金花茶组叶绿体基因组(丁祥青等，2022)和其他被子植物中均有发现(Yang et al., 2014; Hui et al, 2014)，这可能归因于 polyA 和 polyT 相比 polyC 和 polyG 具有更高的结构稳定性(Jin et al., 2023)。然而，我们检测到的 SSR 位点数量与 Hui 等(2014)的结果存在差异，这可能是由于检索 SSR 位点的参数设定不同，导致输出的结果不同。在越南小花金花茶的叶绿体基因组中，cp-SSRs 含量丰富，可用于检测群体的遗传多态性。此外，在越南小花金花茶叶绿体基因组中鉴定出的散在重复序列主要以正向和回文重复为主，这与抱茎金花茶叶绿体基因组序列特征类似(丁祥青等，2022)。这些重复序列是重要的遗传资源，在系统发育研究中发挥着重要作用(Wei et al., 2022)。

密码子使用偏好性对于研究基因的分子进化和外源表达具有重要意义(Li et al., 2021)。本研究首次系统分析了越南小花金花茶叶绿体基因组的密码子使用模式，发现精氨酸(Arg)、异亮氨酸(Leu)、甘氨酸(Ser)是密码子编码的最丰富的氨基酸，由六个密码子编码，相比之下，色氨酸(Trp)、甲硫氨酸(Met)仅有一个密码子编码，叶绿体基因倾向于使用 A / U 密码子，这与耿晓姗等(2022)对金花茶的密码子分析研究结果一致。通过进行中性绘图分析，我们发现 GC12 和 GC3 之间之间没有显著的相关性，且基因分布相对集中，表明自然选择是影响密码子偏好的主要因素(Zhang et al., 2012)。进一步结合 ENC 绘图、PR2-plot 分析结果，表明越南小花金花茶的密码子使用偏好性受多个因素影响，包括突变压力、碱基组成和基因长度，其中自然选择是主导影响因素，这与丁祥青等(2023)的分析结果一致(李清等，2022)。自然选择是叶绿体基因组进化的主要驱动力，这一发现加深了我们对越南小花金花茶进化历史的理解，特别是与自然选择有关的进化历史。对密码子偏好性的分析结果有利于密码子优化，可为今后金花茶组植物转基因技术提供理论依据(段义忠和张凯，2020)。

在叶绿体基因组中，IR 的收缩和扩张时常发生，这些差异会导致假基因的产生、基因重复和基因缺失，进一步造成 IR / SC 连接处的位置变化，高等植物叶绿体基因组长度变异就是由这种位置变化造成的(Li et al., 2021)。本研究对越南小花金花茶及 6 个近缘金花茶组植物的叶绿体基因组边界进行比较分析，结果表明 7 个物种 IR 区长度(25 996 bp ~ 26 096 bp)基本一致，未检测到基因丢失，这表明 IR 区的高度保守可能对于维持其长度和结构稳定

至关重要。通过边界收缩和扩张分析，发现了 *ycf*1 基因的长度和位置在不同金花茶组植物间存在差异，可能是潜在的突变热点区域。在先前的研究中，*ycf*1 基因因其高度的多态性而被推荐作为植物的核心 DNA 条形码，然而，这一潜在的高变区是否可以作为金花茶组植物有效的 DNA 条形码还需要进一步的验证（Li et al., 2021）。此外，越南小花金花茶及 6 个近缘金花茶组植物的叶绿体基因组边界相对保守，表明其亲缘关系较近，之后的系统发育研究也支持了这一推论。

迄今为止，许多学者已经采用诸如随机扩增多态性 DNA (RAPD)、叶绿体 DNA *trn*L-*trn*F、扩增片段长度多态性(AFLP)、简单重复序列间区(ISSR)以及核糖体内转录间隔区(nrITS)等多样化的分子标记，以阐明山茶属之间的亲缘关系(Ju et al., 2021)。 尽管 Wei 等（2022）整合了 ddRAD、转录组、nrITS 和 SSC 等分子标记研究金花茶组系统发育关系，但基于叶绿体基因组 SSC 区的系统发育树支持率极低。本研究基于叶绿体全基因组序列，对越南小花金花茶和其余 23 种金花茶组植物的系统发育关系进行了分析，依据系统发育的结果可将系统发育树分为两个大的分支，其中越南小花金花茶与小花金花茶聚为高支持率的分支，亲缘关系最近。各分支支持率较高，表明完整的叶绿体基因组数据可为重建金花茶组系统发育关系提供重要的数据支持。

综上所述，该文首次完成了越南小花金花茶叶绿体基因组的序列测序、组装注释和基础信息分析，揭示了越南小花金花茶叶绿体基因组基本特征。在此基础上，通过剖析密码子使用模式以及影响密码子偏好的各类因素，确定了越南小花金花茶叶绿体基因组的高频密码子，为后续山茶属金花茶组植物的转基因研究提供了理论支持。此外，基于全叶绿体基因组构建的系统发育树明确了越南小花金花茶在金花茶组植物中的系统位置。本研究为后续越南小花金花茶和其他金花茶组植物的保护和合理开发利用研究提供了理论基础。

参考文献:

AII J, KISHIMA Y, MIKAMI T et al., 1997. Expansion of the IR in the chloroplast genomes of buckwheat species is due to incorporation of an SSC sequence that could be mediated by an inversion[J]. Curr Genet, 31: 276-279.

AMIRYOUSEFI A, HYVÖNEN J, POCZAI P., 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes[J]. Bioinformatics, 34(17): 3030-3031.

BEIER S, THIEL T, MüNCH T, et al., 2017. MISA-web: a web server for microsatellite prediction[J]. Bioinformatics, 33(16): 2583-2585.

CAVALIER ST, 2002. Chloroplast evolution:secondary symbiogenesis andmultiple losses[J]. Curr Biol, 12: 62-64.

CHAKRABORTY S, YENGKHOM S, UDDIN A, 2020. Analysis of codon usage bias of chloroplast genes in *Oryzaspecies*[J]. Planta, 252: 1-20.

丁祥青，毕远洋，陈佳婷，等，2022. 抱茎金花茶(*Camellia tienii*)的叶绿体基因组特征分析[J]. 江苏农业科学，50(23): 33-40. [DING XQ, BI YY, CHEN JT, et al., 2022. Analysis of the chloroplast genome characteristics of *Camellia tienii*[J]. Agric Sci Jiangsu, 50(23): 33-40.]

丁祥青，陈丝雨，陈佳婷，等，2023. 11 种金花茶叶绿体基因组密码子偏好性分析[J]. 福建农林大学学报(自然科学版)，52(4): 1-9. [DING XQ, CHEN SY, CHEN JT, et al., 2023. Codon bias analysis of 11 yellow Camellia chloroplast genome[J]. J Fujian Agric For Univ(Nat Sci Ed), 52(4): 1-9.]

丁祥青，李文芳，吴丽君，等，2022. 4 种金花茶叶绿体基因组的比较分析[J]. 福建农林大学学报(自然科学版)，52(03): 1-11. [DING XQ, LI WF, WU JL, et al., 2022. Chloroplast genome characteristics and genetic relationship of yellow *Camellia*[J]. J Fujian Agric For

Univ (Nat Sci Ed), 52(03): 1-11.]
DONG W, XU C, WU P, et al., 2018. Resolving thesystematic positions of enigmatic taxa: Manipulating the chloroplast genome data of *Saxifragales*[J]. Mol Phylogenet Evol, 126: 321-330.
段义忠，张凯，2020. 沙冬青属植物叶绿体基因组对比和系统发育分析[J]. 西北植物学报，40(8): 1323-1332. [DUAN YZ, ZHANG K, 2020. Comparative analysis and phylogenetic evolution of the complete chloroplast genome of *Ammopiptanthus*[J]. Acta Bot Boreal-Occident Sin, 40(8): 1323-1332.]
耿晓姗，贾魏，陈佳宁，等，2022. 金花茶叶绿体基因组密码子偏好性分析[J]. 分子植物育种，20(7): 2196-2203. [GENG XS, JIA W. CHEN JN, et al., 2022. Codon usage bias analysis of chloroplast genome in *Camellia nitidissima* [J]. Mol Plant Breed, 20(7): 2196-2203.]
GEORGE O, ANTHONY SC, 2015. In pursuit of hidden camellias 32 new *camellia* species from Vietnam and China [M]. 2nd edition. Sydney: Theaceae Exploration Associates.
GUI L, JIANG S, XIE D, et al., 2020. Analysis of complete chloroplast genomes of curcuma and the contribution to phylogeny and adaptive evolution[J]. Gene, 732:144355.
GU J, CHEN S, ZHOU Y, et al., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor[J]. Bioinformatics, 34(17): i884-i890.
HENRY D, LIN CS, YU M, et al., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering[J]. Genome Biol, 17(134): 1-29.
HU XY, XU YQ, HAN YZ, et al., 2019. Codon usage bias analysis of the chloroplast genome of *Ziziphus jujuba* var. spinosa[J]. J For Environ, 39(6): 621-628.[胡晓艳，许艳秋，韩有志，等，2019. 酸枣叶绿体基因组密码子使用偏性分析[J]. 森林与环境学报，39(6): 621-628.]
HUI H, CHAO S, YUAN L, et al., 2014. Thirteen *camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships[J]. BMC Evol Biol, 14(1): 151-151.
JIAN JJ, YU WB, YANG JB, et al., 2018. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes[J]. Genome Biol, 21(1): 241.
JIN GZ, LI WJ, SONG F, et al., 2023. Comparative analysis of complete Artemisia subgenus Seriphidium (Asteraceae: Anthemideae) chloroplast genomes: insights into structural divergence and phylogenetic relationships[J]. BMC Plant Biol, 136: l-23.
JU NG, HOANG DKD, KIM CK, et al., 2021. Complete chloroplast genomes shed light on phylogenetic relationships, divergence time, and biogeography of Allioideae (Amaryllidaceae)[J]. Sci Rep, 11(1): 3262
KATOH K, STANDLEY DM, 2013. MAFFT mutiple sequence alignment software version 7:improvements in performance and usabilit[J]. Mol Biol Evol, 30(4): 772-780.
KURTZ JV, CHOUDHURI, OHLEBUSCH E, et al., 2001. REPuter: the manifold applications of repeat analysis on a genomic scale[J]. Nucleic Acids Res, 29(22): 4633-4642.
LI DM, LI J, WANG DR, et al., 2021. Molecular evolution of chloroplast genomes in subfamily *Zingiberoideae* (Zingiberaceae)[J]. BMC Plant Biol, 21: 1-24.
LI GE, JIANG CJ, QI Y, et al., 2022. Morphological Characteristics and identification points of *Camellia minima* and *Camellia cucphuongensis* from Vietnam[J]. Southern Hortic, 33(1): 54-60. [李桂娥，蒋昌杰，漆娅，等，2022. 越南小花金花茶和菊芳金花茶的形态特征及识别要点[J]. 南方园艺，33(1): 54-60.]
LI L, HU YF, HE M, et al., 2021. Comparative chloroplast genomes:insightsinto the evolution of

the chloroplastgenome of *Camellia sinensis* and thephylogeny of Camellia[J]. BMC Genomics, 22(138).

LI Q, LUO RJ, GE R, et al., 2022. Analysis on Codon Usage Bias of Chloroplast Genomein *Ampelopsis grossedentata*[J]. Guangdong Agr Sci, 49(11): 162-169.

LI W, ZHANG CP, GUO XP, et al., 2019. Complete chloroplast genome of *Camelliajaponica* genome structures, comparative andphylogenetic analysis[J]. PLoS ONE, 14(5): e0216645.

LI XP, MENG J, ZHANG N, et al., 2018. Comparative analysis of chloroplast genomes of *Aconitum vilmorinianum* and *Aconitum vilmorinianum* var. *patentipilum*[J]. J Chin Med Mat, 41(8): 1812-1820. [李雪佩，孟静，张琳娜，等，2018. 黄草乌与展毛黄草乌叶绿体全基因组结构的比较分析[J]. 中药材，41(8): 1812-1820.]

LIU H, HUANG Y, DU X, et al., 2012. Patterns of synonymous codon usage bias in the model grass *Brachypodium distachyon*[J]. Genet Mol Res, 11(4): 4695-4706.

LIU Q, LI Y, YANG RM, et al., 2021. Yellow *Camellia*: Resource Status and Research Progress in Modern Studies[J]. Mod Chin Med, 23(04): 727-733. [刘青，李月，杨润梅，等，2021. 金花茶组植物资源现状与现代研究进展[J]. 中国现代中药，23(4): 727-733.]

LIU S, NI Y, LI J, et al., 2023. CPGView: A package for visualizing detailed chloroplast genome structures[J]. Mol Eco Resourc, 23(3): 694-704.

MINH BQ, SCHMIDT HA, CHERNOMOR O, et al., 2021. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era[J]. Mol Biol Evol, 37(5): 1530-1534.

National Forestry and Grassland Administration of PRC, 2021. List of National Key Protected Wild Plants [EB]. 2021(15). [中华人民共和国国家林业和草原局，2021. 国家重点保护野生植物名录. [EB] 2021 年第 15 号.]

PAN S, MOU C, WU H, et al., 2020. Phylogenetic and codon usage analysis of atypical porcine pestivirus (APPV)[J]. Virulence, 11(1): 916-926.

PASCAL L, STEPHAN G, 2018. GB2sequin-A file converter preparing custom GenBank files for database submission[J]. Genomics, 111(4): 759-761.

QI YY, XU WJ, XING T, et al., 2015. Synonymous codon usage bias in the plastid genome is unrelated to Gene structure and shows evolutionary heterogeneity[J]. Evol Bio Online, 11: 65-77.

SAI X, 2018. Anti-lung Cancer Effect of *Camellia euphlbia* Flowers Extract and itsPotential Mechanism of Action[D]. Dalian: Dalian University of Technology. [赛璇，2018. 金花茶花提取物的肺癌防治作用及其机理初探[D]. 大连：大连理工大学.]

SHI L, CHEN H, JIANG M, et al., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer[J]. Nucleic Acids Res, 7(W1): W65-W73.

SHI F C., 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp[J]. iMeta, 2: e107.

SOPHIARANI Y, ARIF U, SUPRIYO C, 2019. Deciphering codon usage patterns and evolutionary forces in chloroplast genes of *Camellia sinensis* var. *assamica* and *Camellia sinensis* var. *sinensis* in comparison to *Camellia pubicosta*[J]. J Inte Agric, 18(12): 2771-2785.

SUEOKA N, 1988. Directional mutation pressure and neutral molecular evolution[J]. Proc Natl Acad Sci USA, 85(8): 2653-2657.

WEI L, HE J, JIA X, et al., 2014. Analysis of codon usage bias of mitochondrial genome in

Bombyx mori and its relation to evolution[J]. BMC Evol Biol, 14: 262.

WEI SJ, LIUFU YQ, ZHENG HE, et al., 2022. Using phylogenomics to untangle the taxonomic incongruence of yellow‐flowered *Camellia* species (Theaceae) in China[J]. J Sys Evol, 00(00): 1-16.

WU LJ, ZHENG HC, CHEN WR, et al., 2020. Performance and Thinking on Introduction and Cultivation of *Camellia* Sect.*Chrysantha* Chang in Fujian[J]. Fujian For Sci Technol, 47(2): 109-115. [吴丽君，郑惠成，陈文荣，等，2020. 福建金花茶组植物引种栽培现状与思考[J]. 福建林业科技，47(2): 109-115.]

WU CS, CHAW SM, HUANG YY, 2013. Chloroplast phylogenomics indicates that *Ginkgo biloba* is sister to *cycads*[J]. Gen Bio and Evol, 5(2): 243-254.

XIANG H, ZHANG RZ, BUTLER RR, et al., 2015. Comparative analysis of codon usage bias patterns in Microsporidian genomes[J]. PLoS ONE, 10: e0129223.

XIN YX, LI RZ, LI X, et al., 2021. Analysis on codon usage bias of chloroplast genome in *Mangifera indica*[J]. J Centr S Univ For Technol, 41(9): 145-157.[辛雅萱，黎若竹，李鑫，等，2021. 杧果叶绿体基因组密码子使用偏好性分析[J]. 中南林业科技大学学报，41(9): 145-157.]

YANG LC, DENG SX, ZHU YQ, et al., 2023. Comparative chloroplast genomics of 34 species in subtribe Swertiinae (Gentianaceae) with implications for its phylogeny[J]. BMC Plant Bio, 23(164): 0-20.

YANG XY, CAI YB, TAN QL et al., 2021. Analysis of Codon Usage Bias in the Chloroplast Genome of *Ananas comosus*[J]. J Trop Crop, 43(3): 439-446[杨祥燕，蔡元保，谭秦亮，等，2021. 菠萝叶绿体基因组密码子偏好性分析[J]. 热带作物学报，43(3): 439-446.]

ZHANG L, NI S, LI JY, et al., 2019. Analysis of petal nutrition and bioactive components in different periods of *Camellia nitidissima*[J]. For Res, 32(2): 32-38.[张蕾，倪穗，李纪元，范正琪，等，2019. 金花茶不同时期花瓣营养与生物活性成分分析[J]. 林业科学研究，32(2): 32-38.]

ZHANG JW, 2019. The complete chloroplast genome and phylogenetic analysis of the endangered species *Syringa pinnatifolia* (Oleaceae)[D]. Xianyang: Northwest A & F University. [张靖雯，2019. 濒危植物羽叶丁香叶绿体全基因组及系统发育研究[D]. 杨凌：西北农林科技大学.]

ZHANG TW, FANG YJ, WANG XM, et al., 2012. The Complete Chloroplast and Mitochondrial Genome Sequences of *Boea hygrometrica*: Insights into the Evolution of Plant Organellar Genomes[J]. PLoS ONE, 7(1): e30531.

ZHANG W, ZHAO Y, YANG G, et al., 2019. Determination of the evolutionary pressure on *Camellia oleifera* on Hainan Island using the complete chloroplast genome sequence[J]. PeerJ, 7: e7210.

ZHANG Y, NIE X, JIA X, et al., 2012. Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family[J]. Aust J Bot, 60: 461-470.

ZHAO YC, ZHENG H, XU AX, et al., 2016. Analysis of codon usage bias of envelope glycoprotein genes in nuclear polyhedrosis virus (NPV) and its relation to evolution[J]. BMC Genomics, 17: 677-677.

ZHU WY, LONG Y, ZHENG S, et al., 2022. Chloroplast genome structure and characterization of *Melaleuca bracteata*[J]. Mol Plant Breed, 1-14. [朱伟垚，龙宇，郑叔召，等，2022. 黄金宝树叶绿体基因组测序与特征分析[J]. 分子植物育种，1-14.]